A novel multi-stage residual feature fusion network for detection of COVID-19 in chest X-ray images.

FANG, Z., REN, J., MACLELLAN, C., LI, H., ZHOA, H., HUSSAIN, A. and FORTINO, G.

2021

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMBMC.2021.3099367, IEEE Transactions on Molecular, Biological and Multi-Scale Communications

A Novel Multi-stage Residual Feature Fusion Network for Detection of COVID-19 in Chest X-ray Images

Zhenyu Fang, Jinchang Ren, Calum MacLellan, Huihui Li, Huimin Zhao, Amir Hussain, and Giancarlo Fortino

***Abstract*—To suppress the spread of COVID-19, accurate diagnosis at an early stage is crucial, chest screening with radiography imaging plays an important role in addition to the real-time reverse transcriptase polymerase chain reaction (RT-PCR) swab test. Due to the limited data, existing models suffer from incapable feature extraction and poor network convergence and optimization. Accordingly, a multi-stage residual network, MSRCovXNet, is proposed for effective detection of COVID-19 from chest x-ray (CXR) images. As a shallow yet effective classifier with the ResNet-18 as the feature extractor, MSRCovXNet is optimized by fusing two proposed feature enhancement modules (FEM), i.e. low-level and high-level feature maps (LLFMs and HLFMs), which contain respectively more local information and rich semantic information, respectively. For effective fusion of these two features, a single-stage FEM (MSFEM) and a multi-stage FEM (MSFEM) are proposed to enhance the semantic feature representation of the LLFMs and the local feature representation of the HLFMs, respectively. Without ensembling other deep learning models, our MSRCovXNet has a precision of 98.9% and a recall of 94% in detection of COVID-19, which outperforms several state-of-the-art models. When evaluated on the COVIDGR dataset, an average accuracy of 82.2% is achieved, leading other methods by at least 1.2%.**

***Index Terms*—COVID-19, chest x-ray imaging, MSRCovXNet, feature enhancement module, ResNet-18**

I. Introduction

ON January 30, 2020, the World Health Organization (WHO) formally announced the novel coronavirus pneumonia (COVID-19) as a global health emergency [1] , and from March 31, 2020, this was declared as a pandemic [2]. With millions of infected cases and deaths reported in the world [3], COVID-19 has rapidly spread to hundreds of countries and regions. As reported in [4], [5], it has caused more deaths, than the previous coronavirus strains, for instance, the Middle East Respiratory Syndrome (MERS) and the Severe Acute Respiratory Syndrome (SARS). By the end of 2020, the CVOID-19 pandemic has taken massive losses, with respect to the population health [6] and economic recession [7], from many countries. Apart from the prediction model of epidemiological trends [8], it becomes crucial to develop useful tools for fast and effective diagnoses and triaging of patients with suspected COVID-19 symptoms.

Currently, there are two ways in diagnosing of COVID-19, i.e. polymerase chain reaction (RT-PCR) swab test [9] and chest radiography imaging (CRI). RT-PCR swab test, which detects the viral RNA from sputum or nasopharyngeal swab, is now most popularly used for diagnosing COVID-19. However, it may introduce false detections or missing detections, regardless a lengthy waiting time for the results to be released. Studies in [10] have found that false negative rate of the RT-PCR swab test is high, which requires repeated tests for a reliable diagnosis.

As a useful supplementary to the RT-PCR swab test, CRI based diagnostic diagnoses the patients with suspected COVID-19 symptoms through visual analysis of the thoracic lesions on the computed tomography (CT) or chest X-ray (CXR) screening [11]. Compared with CXR imaging, CT is found to be more suitable for COVID-19 detection [12]. As CT imaging is generally more expensive and time-consuming, CXR imaging is thus more popularly used in detecting COVID-19, though with a loss of image resolution and contrast [13]. For both CRI methods, however, key radiological features found in COVID-19 cases, including ground-glass opacities, bilateral involvement, peripheral distributions and crazy-paving patterns, are also partially presented in MERS and SARS [14]. Furthermore, clinical symptoms of COVID-19, such as fever and cough, are similar to viral pneumonia [14]. With a limited period to gain relevant experience, it is a challenging task for the radiologists to discriminate COVID-19 from other pneumonias. With the increased cases of infections, the pressure on health services keeps rising. It is therefore essential to develop a robust and effective computer-aided diagnosis systems to reduce the diagnostic period and alleviate the burden on the clinical staff.

In recent years, deep convolutional neural networks (DCNN) are validated as an effective tool for multiple medical image processing tasks, such as classification, lesion segmentation, and reconstruction. Before DCNN, traditional machine learning (ML) models detect diseases based on extraction of hand-crafted features, which is time-consuming and lack of generalizability. Surpassing over traditional ML approaches, DCNN enables automatic feature extraction during the training, hence it is more efficient on feature searching and more robust on testing on the new data. For the sake of robustness

Z. Fang, J. Ren, H. Li and H. Zhao are with School of Computer Sciences, Guangdong Polytechnic Normal University, Guangzhou, China

Z. Fang is also with School of Computer Software and Microelectronics, Northwestern Polytechnical University, Xi'an, China.

J. Ren is also with the National Subsea Centre, Robert Gordon University, Aberdeen, corresponding author, E-mail: jinchang.ren@ieee.org.

C. MacLellan is with Centre for Signal and Image Processing, University of Strathclyde, Glasgow, UK.

A. Hussain is with School of Computing, Edinburgh Napier University, Edinburgh, U.K.

G. Fortino is with Dept. of Informatics, Modelling, Electronics and Systems, University of Calabria, Italy.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMBMC.2021.3099367, IEEE Transactions on Molecular, Biological and Multi-Scale Communications

and efficiency, DCNN has been successfully applied in many tasks of medical image analysis, such as detecting retinal diseases [15], breast cancer lesions [16], and brain tumours [17]. For the applications in terms of thoracic imaging, one study [18] has empirically validated that the DCNN can outperform experienced radiologists on classification of 14 thoracic diseases. In the context of COVID-19 diagnosis, existing DCNN based methods can also well address this challenge by extending the depth of network [19] or adopting the model assembling [20].

Although increasing the number of layers (i.e. using a deeper network) can improve the capability of feature extraction, this requires the dataset to be sufficiently large (e.g. millions of images, the similar scale as the ImageNet [21]). Due to the limited availability of the COVID-19 data, the efficacy of the existing deep learning models is severely affected, resulting in less capable feature extraction and difficulty of network convergence and optimization. To tackle these issues, in this paper, a multi-stage residual network, MSRCovXNet, is proposed for effective detection of COVID-19 from the CXR images. We aim to derive highly discriminative features from a shallow network, where the number of samples are limited. To achieve this, a ResNet-18 [22] is used as the feature extractor, which is optimized by the fusion of features from multiple stages for improved classification and decision-making.

The major contributions of this paper can be summarized as follows:

i. Taking the ResNet-18 as the feature extractor, a shallow yet effective COVID-19 classifier, MSRCovXNet, is proposed for effective detection of COVID-19 under limited training samples;
ii. A single stage feature enhancement module (SSFEM) is proposed to enhance the feature representation of low-level features, whilst a multi-stage feature enhancement module (MSFEM) is proposed to obtain highly discriminative features fused from multiple stages;
iii. Without ensembling other deep learning models, the proposed MSRCovXNet has a precision of 98.9% and a recall of 94% in the detection of COVID-19 cases, achieving state-of-the-art performance on the COVIDx dataset. When evaluated on the COVIDGR dataset, an average accuracy of 82.2% is achieved, leading other methods by at least 1.2%. When compared with other CNN models trained on different datasets, the proposed method still shows superior performance.

The remaining parts of this paper are organized as follows. Section II briefly introduces the related work. The architecture of the proposed method and the experimental results are detailed in Section III and Section IV, respectively. Finally, some concluding remarks are given in Section V.

II. Related work

Since the outbreak of the COVID-19, a number of DCNN-based models have been developed for the detection of COVID-19 from CT and CXR images. At first, many people focused on a two-category classification, in which the COVID-19 cases were distinguished from either healthy cases [31], [32], or other lung infections diseases, such as viral pneumonia [33], [34], [35], [36] and others [37], [38], [39], [40]. Most of these methods report seemingly impressive results, where performances in the range of 90-100% are not uncommon. However, since doctors not only need to determine whether their patient has COVID-19 or not, but to also identify whether a patient with suspected COVID-19 symptoms does indeed have COVID-19 or a similarly presented infection, the two-class approaches over-simplify the detection problem. Without taking into account the possibility of patients having a healthy image, or being unable to distinguish between various pneumonias, the proposed models most likely encourage a greater degree of overfitting to the training data. As a result, there has been an increased trend in the literature towards adopting a three-class approach, where models are trained to detect healthy patients, as well as discriminating between images of COVID-19 from other pneumonias[1]. This will improve the model's diagnostic sensitivity, and additionally help doctors have a better understanding of what separates COVID-19 images from other pneumonias presenting similar features.

For this reason, our work strives to contribute to the body of work addressing the three-class problem for detecting COVID-19 from CXR images, where many methods have been proposed. In Wang *et al.* [41], a deep CNN model, namely COVID-Net, is proposed, which is actually one of the first three-class deep learning models on CXR diagnosis. The three categories as defined in COVIDx dataset [41], [42] include COVID-19, pneumonia, and healthy cases, respectively. For performance assessment, the COVIDx dataset is collected, which includes 8066, 5551 and 386 normal, bacterial pneumonia (containing both viral pneumonia and bacterial pneumonia), and COVID-19 patients, respectively. A F1 score of 95.9% has been reported on the testing set.

In comparison to a single model used in Wang *et al.* [41], Karim *et al.* [20] have proposed an ensemble of DenseNet-161 and VGG-19 and form the DeepCOVIDExplainer model. Experiments on a dataset with 11,896 images in total have achieved a precision of 89.61% and a recall of 83% on 77 COVID-19 test samples, where the categories are the same as in Wang *et al.*. In order to provide an interpretable evidence to the clinical staff, the class-discriminating pixels on the test images are visualized using the Grad-CAM++ method [43].

To cope with the high degree of imbalance within the categories of the collected COVID-19 samples, Bassi *et al.* [23] have applied the transfer learning to pretrain the proposed CheXNet [18] on the dataset of 112,120 CXR images with 14 thoracic diseases, including the pneumonia samples. For easy adopting of the pretrained model on the target dataset, the output of the last fully-connected layer is reduced to 3 to coping with the three categories of cases. The dataset used for training in [23] contains 127, 1285 and 1281 COVID-19, pneumonia, and normal CXR images, respectively. With the help of data augmentation, an average classification accuracy of 97.8% on a testing set of 180 images is achieved.

[1]In this case, viral and bacterial pneumonias are grouped into a single 'pneumonia' class for comparing with images of the COVID-19 and healthy patients to give the three classes.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMBMC.2021.3099367, IEEE Transactions on Molecular, Biological and Multi-Scale Communications

TABLE I: Summary of methods and findings for three-class (normal, pneumonia, and COVID-19) approaches on Chest X-ray images.

Reference	Dataset	Method	Results
Wang [19]	COVID-19 386 Pneum. 5551 Normal 8066	CNN	95.9% (F1 score)
Karim [20]	COVID-19 259 Pneum. 8614 Normal 8066	DenseNet, ResNet, VGG19 (ensemble)	89.1% (Prec.) 83.0% (Rec.)
Bassi [23]	COVID-19 219 Vir. Pneum. 1345 Normal 1341	Pre-trained CheXNet (DenseNet-121)	98.3% (Acc.)
Zhang [24]	Training / Testing COVID-19 258 / 60 Pneum. 2306 / n/a Normal 8154 / 885	Semi-supervised domain adaption.	93.0% F1 score
Kim [25]	COVID-19 184 Pneum. 4245 Normal 1579	Ensemble three ResNets	94% (Prec.) 100% (Rec.)
Yamac [26]	COVID-19 462 Bact. Pneum. 2760 Vir. Pneum. 1485 Normal 1579	Pre-trained CheXNet (DenseNet121)	95.6% (Acc) 93.9% (Rec.) 95.8% (Spec.)
Chawki [27]	COVID-19 231 Bact. Pneum. 2780 Vir. Pneum. 1493 Normal 1583	Inception-ResNetV2	92.2% (Acc)
Rahimzadeh [28]	COVID-19 180 Pneum. 6054 Normal 8851	Xception, ResNetV2 (ensemble)	99.5% (Acc. COVID-19) 91.4% (Acc. all classes)
Togacar [29]	COVID-19 295 Pneum. 98 Normal 65	MobileNetv2, SqueezeNet SVM as classifier	100.0% (Acc. COVID-19) 99.3% (Acc. Normal and Pneum.)
Lv [30]	Dataset 1: Bact. Pneum. 2772; Vir. Pneum. 1493; Normal 1591 Dataset 2: COVID-19 125; other viruses 316	Cascade network of SEME-ResNet50 and SEME-DenseNet169.	97.1% (Acc. COVID-19) 85.6% (Acc. Pneum.)

In [24], Zhang *et al.*, domain shift between datasets, namely COVID-DA, is attempted to improve the classification accuracy under a semi-supervised framework, where both labelled and unlabelled data are utilized for learning the model. The training set is composed of 8154, 2306 and 258 normal, pneumonia, and COVID-19 CXR images, respectively. However, the testing set has only two categories, i.e. 885 normal and 60 COVID-19 images, respectively, where an F1-score of 92.98% and AUC of 0.985 are reported.

In [25], similar to Karim *et al.*, a three ResNets were ensemble with each subnet being trained for classifying a single category, dividing the three-category classification task into three binary classification tasks. The three binary classifiers are first trained using normal (1579 cases) vs diseased (4429 cases), pneumonia (4245 cases) vs non-pneumonia (1763 cases), and COVID-19 (184 cases) vs non-COVID-19 (5824 cases), respectively. Afterwards, the three ensembled ResNets are fine-tuned on another dataset with 1579 normal, 4245 pneumonia, and 184 COVID-19 cases, respectively. Eventually, a precision of 94% and a recall of 100% are achieved.

In Chawki *et al.* [27], the pneumonia is split into bacterial pneumonia and viral pneumonia before conducting a comprehensive study on several popularly used models, such as VGG, ResNet, DenseNet, Inception-ResNet, Inception-V3, and MobileNet-V2. The dataset used contains 1583 normal, 2780 bacterial, 1493 viral, and 231 COVID-19 cases, where 80% of the samples are used for training and the remaining 20% for testing. Finally, they have found that the Inception-

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMBMC.2021.3099367, IEEE Transactions on Molecular, Biological and Multi-Scale Communications

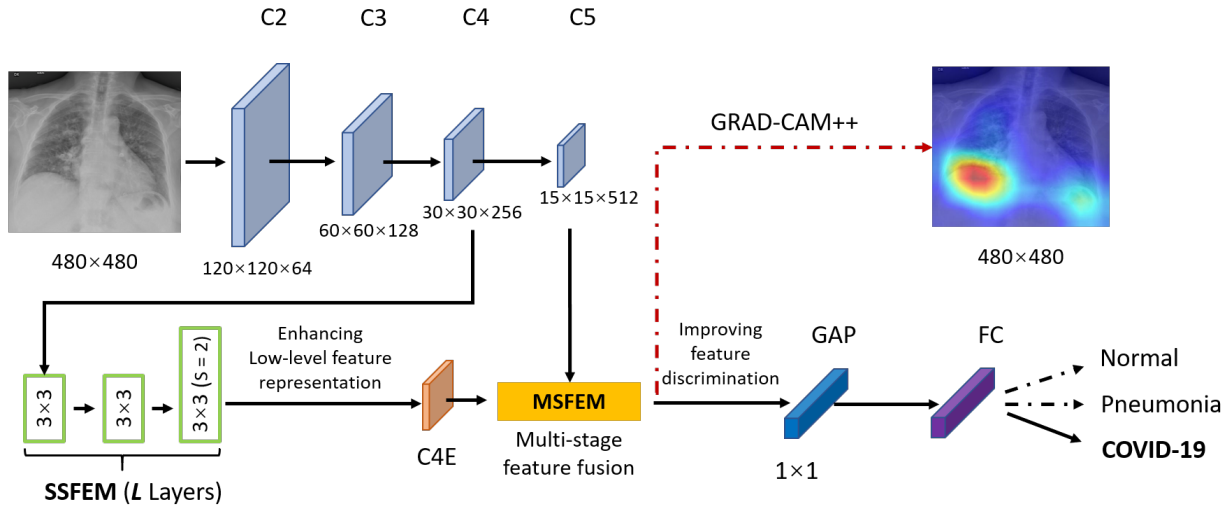


Fig. 1: Architecture of the proposed MSRCovXNet.

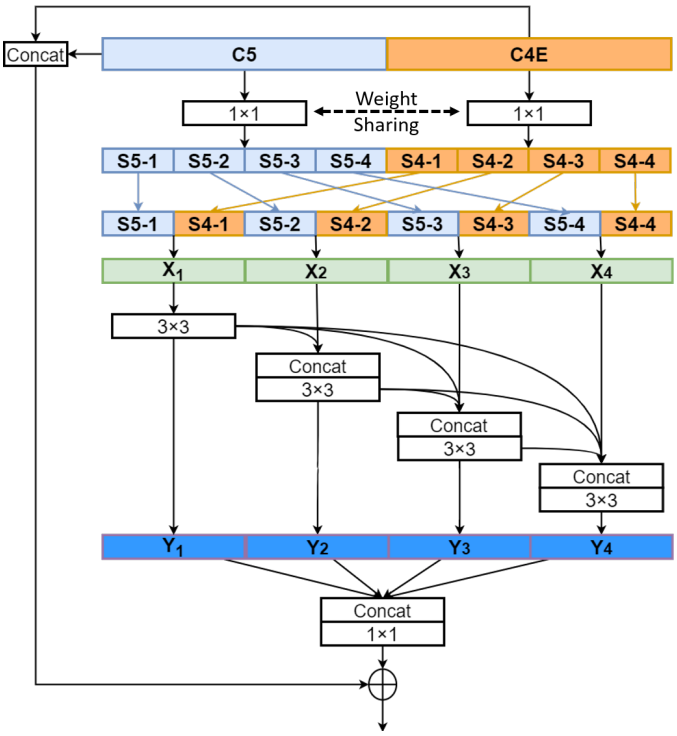


Fig. 2: Architecture of the proposed multi-stage feature enhancement module.

ResNet-V2 can outperform all others with an overall accuracy and F1-score of 92.18% and 92.07%, respectively. In [44], an ensemble of DNNs is proposed for COVID-19 prediction.

In a recent study [45], a quantification method is proposed for quantifying the level of COVID-19 infection severity, which is adopted from a previously defined Radiographic Assessment of Lung Edema (RALE) score [46]. By measuring the extent of the lesion features on each lung, the severity is quantified in a range of 0 to 8, based on which the severity is further defined in 4 levels, i.e. Normal 0, Mild 1-2, Moderate 3-5 and Severe 6-8 [47]. A two-class COVIDGR dataset is collected, containing 426 Normal images and 426 COVID-19 images, in which the positive cases include 76 Normal-PCR+, 100 Mild, 171 Moderate, and 79 Severe cases, respectively. Specifically, Normal-PCR+ indicates that the radiologist does not find any visual lesion regions, although the PCR test is positive. By conducting a 5-fold cross-validation testing with 5 runs on the COVIDGR, an average accuracy of 81.00% can be achieved by the proposed COVID-SDNet [47].

For a more concise comparison of the three-class CXR literature, we have listed the discussed literature along with other relevant papers, together with their respective methods and results, in Table I. There have also been similarly conducted studies implementing deep learning-based classifiers on CT images [48], [49], [50], [51] and lung ultrasonography [52], however we forego discussing them in any detail since our focus is with CXR-based models. A more detailed review of recent CT based deep learning models can be found here [13], [53].

In this paper, we aim to improve the feature representation for a shallow network, which is suitable for training on a small number of samples, without using the ensemble strategy.

III. Proposed Method

In this section, we will discuss the proposed method in details, including, the architecture of the proposed MSRCovXnet, especially the implementation of the multi-stage feature enhancement module, which is proposed for improving the feature representation. After that, the training hyperparameters will also be introduced.

A. Network Architecture

The overall architecture of the proposed MSRCovXNet is shown in Fig. 1. At the current stage, as shown in Table I, there is no large dataset (as the ImageNet) for COVID-19 classification. Under limited samples, a shallow network may perform better than a deep network [22], which is also

verified by the experiments in this paper. Thus, the ResNet-18 [22], which is pretrained on the ImageNet [21], is used as the feature extractor. We denote the output of the blocks "conv2_x", "conv3_x", "conv4_x", and "conv5_x" as "C2", "C3", "C4" and "C5", respectively, where the total stride with respect to these four blocks are 4, 8, 16, and 32. The prediction Y of the original ResNet can be expressed as below

$$Y = fc(\ gap(\ C5\ )\ ); \tag{1}$$

where $fc(\ )$ and $gap(\ )$ are the fully connected layer (FC) and the global average pooling layer (GAP), respectively.

To enhance the capability of feature extraction, feature maps derived from multiple stages, rather than a single stage as the original ResNet-18, are combined in our model. In this paper, we adopt both the C4 and C5 blocks for the lateral classification procedure. Feature maps before C4 are not utilized, simply because the object ("lung" for this task) in the x-ray image is still too large in the previous blocks. As low-level features are sensitive to small objects [54], this affects their impact on the classification task, which is also verified by the experimental results in this paper. For the feature map C4, a single-stage feature enhancement module (SSFEM), which is a subnet with L convolution layers of size 3×3, is assigned to enhance the lower-level feature extraction (denoted by green rectangles in Fig. 1). The size of C4 is reduced, with the number of channels doubled, in the last layer of the subnet to keep the size of the output (denoted as "C4E") the same as C5.

The capability of feature extraction can be further enhanced by using a feature fusion module [55], [56], which is applied to fuse features from different stages. Thus, we present a multi-stage feature enhancement module (MSFEM) to further enhance the extracted feature, which is shown in Fig. 2. The architecture of MSFEM will be detailed in the next subsection. In the end of the network, the enhanced multi-stage feature will be fed to a global average pooling layer, and the final prediction is conducted by a fully connected layer, which is the same as the previous methods [22], [57], [58]. In summary, the prediction Y of the proposed method can be expressed as below:

$$Y = fc(\ gap(\ C45E\ )\ ); \tag{2}$$

where $C45E$ is the feature map extracted using the proposed SSFEM ($ssfem()$) and MSFEM($msfem()$):

$$\begin{aligned} C4E &= ssfem(\ C4\ ); \\ C45E &= msfem(\ concat(\ C4E,\ C5\ )\ ); \end{aligned} \tag{3}$$

where $concat$ is the short of concatenation.

B. Multi-stage Feature Enhancement Module

As shown in Fig. 2, The proposed MSFEM adopts the residual learning [22] for feature fusion and feature refinement, which can be mathematically expressed by:

$$C45E = M(\ C45_{concat}\ ) + C45_{concat}; \tag{4}$$

where $M()$ and $C45E$ are respectively present the convolution layers in the proposed MSFEM and the feature concatenation of $C4$ and $C5$.

First of all, two 1×1 convolution layers are applied to C5 and C4E for adjusting the number of channels, which can be expressed as below:

$$\begin{aligned} C4E_{norm} &= F_{1\times1}(\ C4E\ ); \\ C5_{norm} &= F_{1\times1}(\ C5\ ); \end{aligned} \tag{5}$$

where $F_{1\times1}()$ indicates the two 1×1 convolution layer. As to the number of layers in terms of C4E and C5 are the same, the weights of two 1×1 convolution layers are shared for reducing the number of training parameters. The effectiveness of such implementation is also validated during our experiments: compared with non-weight-sharing, the F1 score of the newtork using weight-sharing could be slightly increased by about 0.1%. As suggested in [58], [59], [60], multi-scale representations, which are acquired in a large range of receptive fields, bring benefits for more accurate prediction. Thus, we utilize multiple subnets, consisting of a variety range of convolutional layers, to achieve multi-scale receptive fields in the proposed MSFEM. However, it will remarkably increase the number of training parameters, if directly taking the fused feature maps as the input of those subnets. As presented above, this will then increase the difficulty of the network optimization, causing the reduction on the detection accuracy. It is also validated in the Section IV, where the proposed outperforms methods with deeper networks. Thus, in the proposed MSFEM, only a part of input channels is fed to each subnet, instead of all channels. To achieve this, after normalizing the number of channels, feature maps from C4E and C5 are divided into N splits. Let $feat_k$ be the k_{th} channel in terms of the input feature map $feat$ ($feat \in \{C4E_{norm}, C5_{norm}\}$). The i_{th} split $S-i$ can be the acquired by:

$$\begin{aligned} S-i = concat(&feat_{1+(i-1)\times(512\backslash\backslash N)}, \\ &feat_{2+(i-1)\times(512\backslash\backslash N)}, \\ &\dots \\ &feat_{(512\backslash\backslash N)+(i-1)\times(512\backslash\backslash N)})\ (i=1,2,3,4) \end{aligned} \tag{6}$$

as suggested in Res2Net [58], we assign $N=4$ as the number of splits and utilize the number of channels with respect to each split as 208 in this paper. Splits from C4E and C5 are concatenated by orders, which can be mathematically expressed as follows:

$$X_i = concat(S4-i,\ S5-i)\ (i=1,2,3,4) \tag{7}$$

where $S4-i$, $S5-i$ are the splits from $C4E_{norm}$ and $C5_{norm}$, respectively; $concat$ is the short of concatenation; X_i is the input split of the following Densely connected block [57].

For each split X_i, the input is the concatenation of all the outputs from the previous layer, as well as the X_i, which can be represented as:

$$Y_i = F(concat(X_i, Y_{i-1}, Y_{i-2}, ..., Y_1))\ (i=2,3,4) \tag{8}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMBMC.2021.3099367, IEEE Transactions on Molecular, Biological and Multi-Scale Communications

TABLE II: Data distribution of the COVIDx dataset

	Normal	Pneumonia	COVID-19	Total
Train	7966	5451	286	13703
Test	100	100	100	300

TABLE III: Selection of feature stages

Blocks	F1 score(%)		
	Normal	**Pneumonia**	**COVID-19**
C5	93.5	93.1	95.3
C5+C4	**94.6**	**94.4**	**95.9**
C5+C4+C3	92.6	94.0	95.9

TABLE IV: F1 scores in terms of the SSFEM layer numbers

Number of L	F1 score(%)		
	Normal	**Pneumonia**	**COVID-19**
1	**94.6**	**94.4**	**95.9**
2	94.1	93.0	95.9
4	93.3	94.4	96.4
6	93.5	94.1	96.4

TABLE V: Ablation study of the MSFEM layer

	F1 score(%)		
	Normal	**Pneumonia**	**COVID-19**
ResNet-18 +SSFEM (c=1)	**94.6**	94.4	95.9
ResNet-18 +SSFEM (c=1) +MSFEM	94.2	**95.4**	**96.4**

where Y_i is the output with respect to X_i, $F(.)$ denotes the convolution layer. Specially, for Y_i, the input is X_i, because there is no output "Y_0" before it. In the end, the splits are re-concatenated, and the number of channels is adjusted via a 1×1 convolution layer. For the ResNet-18 based network, the number of channels are 1024 (C5:512, C4E:512) [22].

IV. Experimental Results

A. Experimental Settings

1) Dataset: Experiments in this paper are conducted on the COVIDx dataset [41], which is currently the largest publicly available dataset with respect to the number of COVID-19 cases. COVIDx dataset consists of 13703 images for training and 300 images for testing, where the COVID-19 samples are collected from more than 266 COVID-19 patients. Images in the COVIDx dataset are labeled in three classes: normal, non-COVID19 infection (pneumonia), and COVID-19. The training and testing sets are randomly divided according to the patient ID, which means for a particular patient, the associated data will be used either for training or testing, hence there is no overlapped data in this context. Details of the sample distribution in terms of each class are shown in Table II. We train our models on the training dataset and evaluate the performance on the testing dataset.

Tabik *et al.* [47] argued that the majority COVID-19 cases in COVIDx dataset is at severe level. It is therefore essential to validate the proposed MSRCovXNet on another dataset for validating the performance in early-stage diagnosis. To this end, we also evaluated the proposed MSRCovXNet on the COVIDGR 1.0 dataset [47]. As a two-class dataset, COVIDGR includes 426 samples in each class. The severity distribution of the positive cases is: 76 Normal-PCR+, 100 Mild, 171 Moderate, and 79 Severe, respectively.

2) Implementation details and evaluation metrics: The proposed method is implemented on PyTorch [63]. The input image is resized to 480×480 for efficiency. Following the settings in [41], [25], [26], [30], we adopt the Adam optimizer [64] for its promising performance on weight optimization. The initial learning rate is set to 1e-4, which is decreased via the cosine annealing schedule [65]. The batch size is set to 60 on 3 GPUs. The network is trained by 22 epochs. Moreover, we adopt the data augmentation methods as suggested in COVID-Net [41] for a fair comparison, which include: intensity shift, translation, rotation, horizontal flip, and random resizing. For

TABLE VI: Results comparison in terms of F1 score (%) with several state-of-the-art deep learning models on the COVIDx test dataset.

Methods	F1 Score(%)		
	Normal	**Pneumonia**	**COVID-19**
ResNet-18* [22]	93.5	93.1	95.3
ResNet-50* [22]	93.3	93.9	95.9
Res2Net-50* [58]	93.7	94.9	**96.4**
ChexNet* [23]	**94.2**	94.9	95.9
COVID-Net [41]	92.5	91.6	95.9
MSRCovXNet (proposed)	**94.2**	**95.4**	**96.4**

* Trained by author with 5 runs

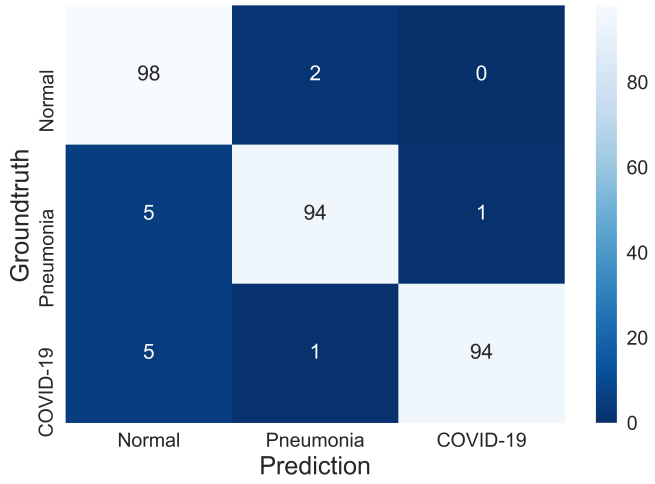


Fig. 3: Confusion matrix for the proposed MSRCovXNet on the COVIDx test dataset. Precision and recall of each class is shown in Table VIII

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMBMC.2021.3099367, IEEE Transactions on Molecular, Biological and Multi-Scale Communications

TABLE VII: Results comparison in terms of F1 score (%) with methods trained on other datasets

Methods	Number of images for training			Number of images for testing			F1 Score (%)		
	Normal	Pneumonia	COVID-19	Normal	Pneumonia	COVID-19	Normal	Pneumonia	COVID-19
Karim [20]	5647	6030	182	2419	2584	77	86.5	84.5	81.6
Zhang [24]	8154	2306	258	885	0	60	n/a	n/a	93.0
Kim * [25]	1422	3821	166	157	424	18	92.2	94.4	96.6
Rahimzadeh [28]	2000	1634	149	6851	4420	31	93.2	89.1	49.1
Togacar ** [29]	46	69	207	19	29	88	**95.3**	**96.5**	**99.5**
MSRCovXNet (proposed)	7966	5451	286	100	100	100	94.2	95.4	96.4
MSRCovXNet-s (proposed)	100	100	100	7966	5451	286	90.0	90.0	80.0

*A slightly higher F1 in COVID-19 cases than ours due to a very small number of test samples in COVID-19 yet the F1 scores for the other two classes are much lower.

**Relatively better performance is due to too limited training and testing samples used, where the higher number of COVID-19 cases than others seems impractical in real scenarios.

TABLE VIII: Precision and recall of the proposed MSRCovXNet

	Normal	Pneumonia	COVID-19
Precision (%)	90.7	96.9	99.0
Recall (%)	98.0	94.0	94.0

performance evaluation, we report the results using the F1 score (%), as it considers both the recall and precision for an overall assessment of false alarms and missed detections.

When training on the COVIDGR dataset, we decrease the training epoch to 15, while all other hyperparameters remain the same as on the COVIDx dataset. Images are cropped to normalize the position of lung as suggested in [47]. Following the same evaluation method in COVIDGR, we conducted 5 different 5-fold cross validations with multiple metrics, including the sensitivity, specificity, precision, F1 and Accuracy. Results of each metric are reported using the average values and the standard deviation over the five runs. As the Normal-PCR+ may impede the overall performance [47], such cases are excluded in our experiments.

B. Ablation Study

In this section, we conduct an ablation study to examine how each proposed component within our MSRCovXNet affects the final performance in the detection of COVID-19.

1) Selection of feature stages: In this section, the selection of feature stages will be discussed. As suggested in the feature pyramid network [66], a single convolution layer is applied as a replacement of the single-stage feature enhancement module (SSFEM). Meanwhile, multi-stage feature enhancement module (MSFEM) is not utilized.

Experimental results are shown in Table III. By fusing the feature maps of C4 and C5, the F1 scores are increased by 0.9%, 1.3% and 0.6% on normal class, pneumonia class and COVID-19 class, respectively. However, after adding the C3, the F1 scores on normal and pneumonia are dropped by 1.2% on average. This indicates that the low-level feature maps (C2 and C3) does not bring benefits to this task. Thus, we only adopt C4 and C5 for feature fusion in the proposed MSRCovXNet.

2) Number of layers in the single-stage feature enhancement module: In this subsection, the number of layers in the SSFEM is discussed. According to the results shown in Table IV, as the number of layers increases, the F1 score of COVID-19 increases by 0.5% on maximum. However, the F1 scores on normal and pneumonia classes decrease. Take $L = 6$ for example, the F1 scores on normal and pneumonia classes are reduced by 0.9% and 0.3%, respectively. This is mainly caused by the size of the dataset. i.e. the complexity of models, where $L > 1$, is overfitting for normal, and pneumonia classes in COVIDx. As the training accuracies are ranged between 97% and 100% for the models in Table IV, we deduce that the performance of SSFEM with deeper layers could be further improved by adding more available samples with more variants in the future. Thus, in this paper, we select $L = 1$ for the following experiments.

3) Effect of the multi-stage feature enhancement module: In this subsection, we will verify the effect of the MSFEM. Experimental results are shown in Table V. As seen, though the F1 score on normal cases decreases by 0.4%, the F1 scores on the pneumonia and COVID-19 classes increase by 1% and 0.5% respectively. Overall, the performance has been further improved with the MSFEM.

C. Compared with the State-of-the-art Methods

1) Result comparison on the COVIDx dataset: Here we compared the proposed MSRCovXNet with the state-of-the-art deep learning models. First of all, the proposed MSRCovXNet

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMBMC.2021.3099367, IEEE Transactions on Molecular, Biological and Multi-Scale Communications

TABLE IX: Result comparison (%) to state-of-the-art methods on the COVIDGR 1.0 dataset. Spec., Prec., and Sens. are the abbreviation of specificity, precision and sensitivity, respectively.

Class	Metric	Methods				
		COVID-Net [41]	COVID-CAPS [61]	FuCiTNetz [62]	COVID-SDNet [47]	MSRCovXNet (proposed)
N	Spec.	83.42 ± 15.39	65.09 ± 10.51	82.63 ± 6.61	$\mathbf{85.20 \pm 5.38}$	82.35 ± 6.49
	Prec.	69.73 ± 10.34	71.72 ± 5.57	79.94 ± 4.28	78.88 ± 3.89	$\mathbf{85.12 \pm 4.07}$
	F1	74.45 ± 8.86	67.52 ± 5.29	81.05 ± 3.44	81.75 ± 2.74	$\mathbf{83.46 \pm 3.13}$
P	Sens.	61.82 ± 22.44	73.31 ± 9.74	78.91 ± 5.88	76.80 ± 6.30	$\mathbf{82.01 \pm 5.76}$
	Prec.	79.50 ± 11.47	68.40 ± 5.13	82.43 ± 5.43	$\mathbf{84.23 \pm 4.59}$	79.73 ± 5.04
	F1	65.64 ± 15.90	70.20 ± 4.31	80.37 ± 3.16	80.07 ± 0.04	$\mathbf{80.60 \pm 2.83}$
Accuracy		72.62 ± 7.6	69.20 ± 3.61	80.77 ± 3.15	81.00 ± 2.87	$\mathbf{82.20 \pm 2.83}$

is compared with the other methods that are trained on the same COVIDx dataset. Experimental results are shown and compared in Table VI, and the confusion matrix is visualized in the Fig. 3. As seen, the proposed MSRCovXNet has achieved state-of-the-art performance in all three classes.

2) Result comparison with methods trained on other datasets: Due to the limited number of image samples, some of the existing methods are trained and evaluated on a subset of the COVIDx dataset [20], [24], [28], [29] or self-collected dataset [23], [25]. However, as the classification task in these datasets are the same, it is worthy to compare the performance with these methods as well. Results are shown in Table VII. Here we compared with methods that used the same classes (normal, pneumonia, and COVID-19) as our methods. Methods with different classes are not compared, e.g. the method in [27] which classified four categories: bacterial pneumonia, coronavirus, COVID-19, and normal.

As seen, the proposed MSRCovXNet outperforms most of the methods in all the three classes, except Togacar [29] and Kim [25]. When compared with Kim, the number of COVID-19 samples is less than 20% of COVIDx. However the F1 score on COVID-19 cases only outperforms by 0.2%, along with a degradation of 2% and 3% on the normal and pneumonia cases, respectively. Therefore, it is hard to say it actually outperformed the proposed MSRCovXNet. For Togacar [29], as the total number of test images is only 136, this is also highly imbalanced for the three classes. As a result, we deduce that the difference on performance is mainly caused by the small size and imbalanced samples of the test set.

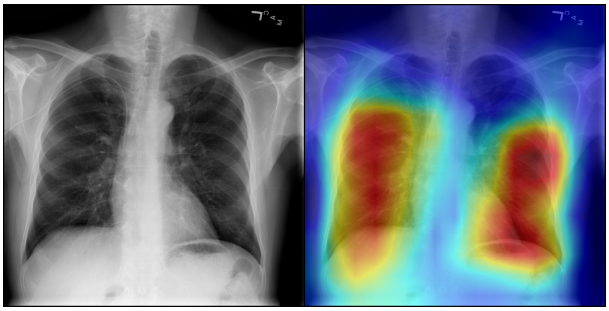
Training on small dataset size. Specifically, we trained the proposed method on the testing set of COVIDx, and test on the training set, which is to further evaluate the efficacy and robustness of the proposed approach in distinguishing COVID-19. In this case, the number of training images are 100 for each class. Results are shown in Table VII as "MSRCovXNet-s". As seen, when the testing set is far larger than the training set, which is similar to the real-life situation, the proposed method can still achieve comparable results to other methods. This validates the effectiveness of the MSRCovXNet on a small training set.

Evaluating on COVIDGR. Results comparison on the COVIDGR dataset from different approaches are listed in Table IX. As seen, even with many early-stage cases included, the proposed MSRCovXNet can still achieve the state-of-the-art performance, surpassing the COVID-SDNet by 1.2% on the average accuracy. This has validated the effectiveness of the proposed method on detecting the early-stage COVID-19. For comparison, although COVID-Net achieves a high F1 score on the COVIDx, the F1 score on the COVIDGR dataset is only 65.64%, which is 15% lower than the proposed method. This has again validated the robustness of the proposed methodology in detection of COVID-19 in CXR images.

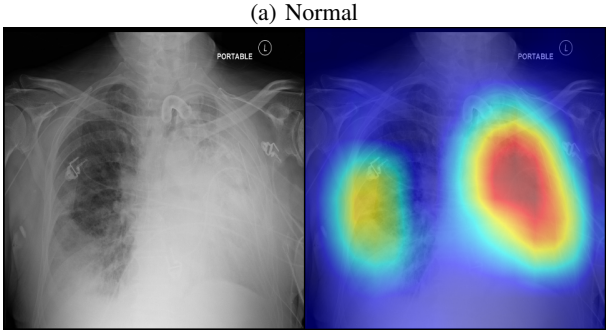
As discussed by Tabik *et al.* [47], the majority of COVID-19 cases in the COVIDx dataset is at the severe level. Methods reported on this dataset have achieved quite high accuracy on detecting COVID-19, due mainly to the low detecting difficulty. This can be also observed in our experiments, see in Table IX, where an accuracy of 82.2% was achieved. However, the classification accuracy drops by 19.27% when the same method is trained and tested on the COVIDGR dataset. This has clearly indicated potential issue of data quality, which may affect the detection accuracy in this context. A lesson herein will be that it is unsuitable to only apply the easy samples for training and testing, where hard samples at less severe levels of COVID-19 would be beneficial. By training on the dataset with uniformly distributed four severity levels, the discrimination of detecting the early-stage cases can be further strengthened.

3) Visualization of the class-discriminating regions: It is important to know the decisive regions on the image where the pixels contribute most to the final decision. This is because it is able to verify the reliability of the diagnostic decision made by the CNN, which can help the clinical doctor to gain a better understanding of the proposed deep learning model. It also benefits them to find out the diseased regions on the image. In this paper, the class-discriminating regions are highlighted using gradient-guided class activation maps (Grad-CAM++) [43], which is shown in Fig. 4. As expected, the proposed method predicts based on regions with pathological features in the lungs, which also validates the high reliability of the proposed method in effective detection of COVID-19 from other cases.

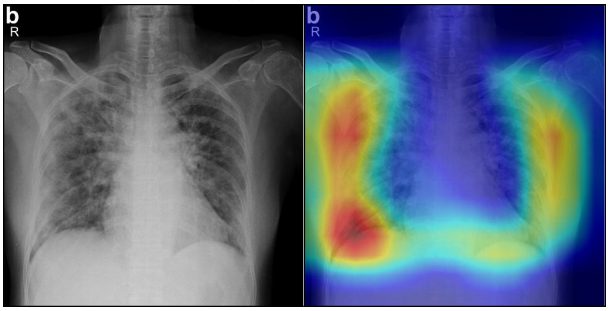
This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMBMC.2021.3099367, IEEE Transactions on Molecular, Biological and Multi-Scale Communications



(a) Normal



(b) Pneumonia



(c) COVID-19

Fig. 4: Decision visualization using Grad-CAM++ with the input label: (a) Normal, (b) Pneumonia, and (c) COVID-19. Images are all from the COVIDx dataset.

V. Conclusion

In this paper, we proposed a novel COVID-19 classifier, namely MSRCovXNet, for the detection of COVID-19 from chest x-ray (CXR) imaging. To tackle the challenging problem of insufficient training samples, the ResNet-18 is used as the feature extractor. In order to improve the discriminative capability of the extracted features from this shallow network, the proposed MSRCovXNet fuses the features from multiple stages rather than adopting the feature map from the last stage for decision making. With a VGG-style subnet structure, the proposed single stage feature enhancement module (SSFEM) has effectively enhanced the feature representation of low-level features. Meanwhile, the proposed multi-stage feature enhancement module (MSFEM) has improved the performance by varying the range of the receptive fields to obtain highly discriminative features fused from multiple stages. The performance of the proposed MSRCovXNet has been validated on the COVIDx dataset, by far the largest publicly available dataset for COVID-19 detection from CXR images. Thanks to the proposed feature enhancement modules, our MSRCovXNet has demonstrated superior performance over several state-of-the-art deep learning models, under a small number of training samples and without any ensembling models. When compared with models which are trained on other datasets, the proposed MSRCovXNet still obtains a promising performance.

Future work will focus on further enhancement of the features using ResNet-style skip connections [22], [67], [58] as the VGG style subnet is suboptimal. In addition, fusion of multiple CXR images will also be utilized as the supplementary information in between can improve the discriminative capability of the learnt features. Finally, we may also explore other deep learning models and also effective methods in addressing imbalanced learning in detection of COVID-19 from CXR and other image data.

Acknowledgment

This work was supported in part by the Dazhi Scholarship of the Guangdong Polytechnic Normal University, the Key Laboratory of the Education Department of Guangdong Province (2019KSYS009), the National Natural Science Foundation of China (62072122, 62006049).

References

[1] W. H. Organization *et al.*, "Statement on the second meeting of the international health regulations (2005) emergency committee regarding the outbreak of novel coronavirus (2019-ncov)," 2005.

[2] ——, "Who director-general's opening remarks at the media briefing on covid-19-11 march 2020," *Geneva, Switzerland*, 2020.

[3] ——, "Coronavirus disease (covid-19): situation report, 135," 2020.

[4] E. Mahase, "Coronavirus: covid-19 has killed more people than sars and mers combined, despite lower case fatality rate," 2020.

[5] C. Wang, P. W. Horby, F. G. Hayden, and G. F. Gao, "A novel coronavirus outbreak of global health concern," *The Lancet*, vol. 395, no. 10223, pp. 470–473, 2020.

[6] M. Chao Hu, M. Yuan Jin, M. Xun Niu, M. Rongyu Ping, and M. Yingzhen Du, "Clinical characteristics of covid-19 patients with digestive symptoms in hubei, china: a descriptive, cross-sectional, multicenter study."

[7] N. Fernandes, "Economic effects of coronavirus outbreak (covid-19) on the world economy," *Available at SSRN 3557504*, 2020.

[8] J. Ren, Y. Yan, H. Zhao, P. Ma, J. Zabalza, Z. Hussain, S. Luo, Q. Dai, S. Zhao, A. Sheikh *et al.*, "A novel intelligent computational approach to model epidemiological trends and assess the impact of non-pharmacological interventions for covid-19," *IEEE Journal Of Biomedical and Health Informatics*, vol. 24, no. 12, pp. 3551–3563, 2020.

[9] M. M. Candace and Daniel, *COVID-19*, 2020 (accessed June, 2020). [Online]. Available: https://radiopaedia.org/articles/covid-19-3?lang=us

[10] J. F.-W. Chan, S. Yuan, K.-H. Kok, K. K.-W. To, H. Chu, J. Yang, F. Xing, J. Liu, C. C.-Y. Yip, R. W.-S. Poon *et al.*, "A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster," *The Lancet*, vol. 395, no. 10223, pp. 514–523, 2020.

[11] C. Huang, Y. Wang, X. Li, L. Ren, J. Zhao, Y. Hu, L. Zhang, G. Fan, J. Xu, X. Gu *et al.*, "Clinical features of patients infected with 2019 novel coronavirus in wuhan, china," *The lancet*, vol. 395, no. 10223, pp. 497–506, 2020.

[12] K. M. Yee, "X-ray may be missing covid cases found with ct," *Korean Journal of Radiology*, 2020.

[13] F. Shi, J. Wang, J. Shi, Z. Wu, Q. Wang, Z. Tang, K. He, Y. Shi, and D. Shen, "Review of artificial intelligence techniques in imaging data acquisition, segmentation and diagnosis for covid-19," *IEEE Reviews in Biomedical Engineering*, 2020.

[14] M. Chung, A. Bernheim, X. Mei, N. Zhang, M. Huang, X. Zeng, J. Cui, W. Xu, Y. Yang, Z. A. Fayad *et al.*, "Ct imaging features of 2019 novel coronavirus (2019-ncov)," *Radiology*, vol. 295, no. 1, pp. 202–207, 2020.

[15] T. Schlegl, S. M. Waldstein, H. Bogunovic, F. Endstraßer, A. Sadeghipour, A.-M. Philip, D. Podkowinski, B. S. Gerendas, G. Langs, and U. Schmidt-Erfurth, "Fully automated detection and quantification of macular fluid in oct using deep learning," *Ophthalmology*, vol. 125, no. 4, pp. 549–558, 2018.

[16] D. A. Ragab, M. Sharkas, S. Marshall, and J. Ren, "Breast cancer detection using deep convolutional neural networks and support vector machines," *PeerJ*, vol. 7, p. e6201, 2019.

[17] K. R. Laukamp, F. Thiele, G. Shakirin, D. Zopfs, A. Faymonville, M. Timmer, D. Maintz, M. Perkuhn, and J. Borggrefe, "Fully automated detection and segmentation of meningiomas using deep learning on routine multiparametric mri," *European radiology*, vol. 29, no. 1, pp. 124–132, 2019.

[18] P. Rajpurkar, J. Irvin, K. Zhu, B. Yang, H. Mehta, T. Duan, D. Ding, A. Bagul, C. Langlotz, K. Shpanskaya *et al.*, "Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning," *arXiv preprint arXiv:1711.05225*, 2017.

[19] C. Wang, P. W. Horby, F. G. Hayden, and G. F. Gao, "A novel coronavirus outbreak of global health concern," *The Lancet*, vol. 395, no. 10223, pp. 470–473, 2020.

[20] M. Karim, T. Döhmen, D. Rebholz-Schuhmann, S. Decker, M. Cochez, O. Beyan *et al.*, "Deepcovidexplainer: Explainable covid-19 predictions based on chest x-ray images," *arXiv preprint arXiv:2004.04582*, 2020.

[21] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE conference on computer vision and pattern recognition*. Ieee, 2009, pp. 248–255.

[22] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[23] P. R. Bassi and R. Attux, "A deep convolutional neural network for covid-19 detection using chest x-rays," *arXiv preprint arXiv:2005.01578*, 2020.

[24] Y. Zhang, S. Niu, Z. Qiu, Y. Wei, P. Zhao, J. Yao, J. Huang, Q. Wu, and M. Tan, "Covid-da: Deep domain adaptation from typical pneumonia to covid-19," *arXiv preprint arXiv:2005.01577*, 2020.

[25] S. Misra, S. Jeon, S. Lee, R. Managuli, and C. Kim, "Multi-channel transfer learning of chest x-ray images for screening of covid-19," *arXiv preprint arXiv:2005.05576*, 2020.

[26] M. Yamac, M. Ahishali, A. Degerli, S. Kiranyaz, M. E. Chowdhury, and M. Gabbouj, "Convolutional sparse support estimator based covid-19 recognition from x-ray images," *arXiv preprint arXiv:2005.04014*, 2020.

[27] K. Elasnaoui and Y. Chawki, "Using x-ray images and deep learning for automated detection of coronavirus disease," *Journal of Biomolecular Structure and Dynamics*, no. just-accepted, pp. 1–22, 2020.

[28] M. Rahimzadeh and A. Attar, "A modified deep convolutional neural network for detecting covid-19 and pneumonia from chest x-ray images based on the concatenation of xception and resnet50v2," *Informatics in Medicine Unlocked*, p. 100360, 2020.

[29] M. Toğaçar, B. Ergen, and Z. Cömert, "Covid-19 detection using deep learning models to exploit social mimic optimization and structured chest x-ray images using fuzzy color and stacking approaches," *Computers in Biology and Medicine*, p. 103805, 2020.

[30] D. Lv, W. Qi, Y. Li, L. Sun, and Y. Wang, "A cascade network for detecting covid-19 using chest x-rays," *arXiv preprint arXiv:2005.01468*, 2020.

[31] O. Gozes, M. Frid-Adar, H. Greenspan, P. D. Browning, H. Zhang, W. Ji, A. Bernheim, and E. Siegel, "Rapid ai development cycle for the coronavirus (covid-19) pandemic: Initial results for automated detection & patient monitoring using deep learning ct image analysis," *arXiv preprint arXiv:2003.05037*, 2020.

[32] A. Sohaib and W. Yi, "Automatic detection of covid-19 using x-ray images with deep convolutional neural networks and machine learning," *medRxiv*, 2020.

[33] S. Wang, B. Kang, J. Ma, X. Zeng, M. Xiao, J. Guo, M. Cai, J. Yang, Y. Li, X. Meng *et al.*, "A deep learning algorithm using ct images to screen for corona virus disease (covid-19)," *MedRxiv*, 2020.

[34] F. Shi, L. Xia, F. Shan, D. Wu, Y. Wei, H. Yuan, H. Jiang, Y. Gao, H. Sui, and D. Shen, "Large-scale screening of covid-19 from community acquired pneumonia using infection size-aware classification," *arXiv preprint arXiv:2003.09860*, 2020.

[35] X. Ouyang, J. Huo, L. Xia, F. Shan, J. Liu, Z. Mo, F. Yan, Z. Ding, Q. Yang, B. Song *et al.*, "Dual-sampling attention network for diagnosis of covid-19 from community acquired pneumonia," *IEEE Transactions on Medical Imaging*, 2020.

[36] H. Kang, L. Xia, F. Yan, Z. Wan, F. Shi, H. Yuan, H. Jiang, D. Wu, H. Sui, C. Zhang *et al.*, "Diagnosis of coronavirus disease 2019 (covid-19) with structured latent multi-view representation learning," *IEEE transactions on medical imaging*, 2020.

[37] B. Ghoshal and A. Tucker, "Estimating uncertainty and interpretability in deep learning for coronavirus (covid-19) detection," *arXiv preprint arXiv:2003.10769*, 2020.

[38] S. Jin, B. Wang, H. Xu, C. Luo, L. Wei, W. Zhao, X. Hou, W. Ma, Z. Xu, Z. Zheng *et al.*, "Ai-assisted ct imaging analysis for covid-19 screening: Building and deploying a medical ai system in four weeks," *medRxiv*, 2020.

[39] C. Jin, W. Chen, Y. Cao, Z. Xu, X. Zhang, L. Deng, C. Zheng, J. Zhou, H. Shi, and J. Feng, "Development and evaluation of an ai system for covid-19 diagnosis," *medRxiv*, 2020.

[40] J. Zhang, Y. Xie, Y. Li, C. Shen, and Y. Xia, "Covid-19 screening on chest x-ray images using deep learning based anomaly detection," *arXiv preprint arXiv:2003.12338*, 2020.

[41] L. Wang and A. Wong, "Covid-net: A tailored deep convolutional neural network design for detection of covid-19 cases from chest x-ray images," *arXiv preprint arXiv:2003.09871*, 2020.

[42] A. Wong, Z. Lin, L. Wang, A. Chung, B. Shen, A. Abbasi, M. Hoshmand-Kochi, and T. Duong, "Towards computer-aided severity assessment via deep neural networks for geographic and opacity extent scoring of sars-cov-2 chest x-rays," *Scientific Reports*, vol. 11, no. 1, pp. 1–8, 2021.

[43] A. Chattopadhay, A. Sarkar, P. Howlader, and V. N. Balasubramanian, "Grad-cam++: Generalized gradient-based visual explanations for deep convolutional networks," in *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*. IEEE, 2018, pp. 839–847.

[44] R. M. Wehbe, J. Sheng, S. Dutta, S. Chai, A. Dravid, S. Barutcu, Y. Wu, D. R. Cantrell, N. Xiao, B. D. Allen *et al.*, "Deepcovid-xr: an artificial intelligence algorithm to detect covid-19 on chest radiographs trained and tested on a large us clinical data set," *Radiology*, vol. 299, no. 1, pp. E167–E176, 2021.

[45] H. Y. F. Wong, H. Y. S. Lam, A. H.-T. Fong, S. T. Leung, T. W.-Y. Chin, C. S. Y. Lo, M. M.-S. Lui, J. C. Y. Lee, K. W.-H. Chiu, T. Chung *et al.*, "Frequency and distribution of chest radiographic findings in covid-19 positive patients," *Radiology*, p. 201160, 2020.

[46] M. A. Warren, Z. Zhao, T. Koyama, J. A. Bastarache, C. M. Shaver, M. W. Semler, T. W. Rice, M. A. Matthay, C. S. Calfee, and L. B. Ware, "Severity scoring of lung oedema on the chest radiograph is associated with clinical outcomes in ards," *Thorax*, vol. 73, no. 9, pp. 840–846, 2018.

[47] S. Tabik, A. Gómez-Ríos, J. L. Martín-Rodríguez, I. Sevillano-García, M. Rey-Area, D. Charte, E. Guirado, J. L. Suárez, J. Luengo, M. A. Valero-González, P. García-Villanova, E. Olmedo-Sánchez, and F. Herrera, "Covidgr dataset and covid-sdnet methodology for predicting covid-19 based on chest x-ray images," *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 12, pp. 3595–3605, 2020.

[48] C. Butt, J. Gill, D. Chun, and B. A. Babu, "Deep learning system to screen coronavirus disease 2019 pneumonia," *Applied Intelligence*, p. 1, 2020.

[49] L. Li, L. Qin, Z. Xu, Y. Yin, X. Wang, B. Kong, J. Bai, Y. Lu, Z. Fang, Q. Song *et al.*, "Artificial intelligence distinguishes covid-19 from community acquired pneumonia on chest ct," *Radiology*, p. 200905, 2020.

[50] T. Javaheri, M. Homayounfar, Z. Amoozgar, R. Reiazi, F. Homayounieh, E. Abbas, A. Laali, A. R. Radmard, M. H. Gharib, S. A. J. Mousavi *et al.*, "Covidctnet: An open-source deep learning approach to identify covid-19 using ct image," *arXiv preprint arXiv:2005.03059*, 2020.

[51] J. Wang, Y. Bao, Y. Wen, H. Lu, H. Luo, Y. Xiang, X. Li, C. Liu, and D. Qian, "Prior-attention residual learning for more discriminative covid-19 screening in ct images," *IEEE Transactions on Medical Imaging*, 2020.

[52] S. Roy, W. Menapace, S. Oei, B. Luijten, E. Fini, C. Saltori, I. Huijben, N. Chennakeshava, F. Mento, A. Sentelli *et al.*, "Deep learning for classification and localization of covid-19 markers in point-of-care lung ultrasound," *IEEE Transactions on Medical Imaging*, 2020.

[53] S. Bhattacharya, P. K. R. Maddikunta, Q.-V. Pham, T. R. Gadekallu, C. L. Chowdhary, M. Alazab, M. J. Piran *et al.*, "Deep learning and medical image processing for coronavirus (covid-19) pandemic: A survey," *Sustainable cities and society*, vol. 65, p. 102589, 2021.

[54] C. Chi, S. Zhang, J. Xing, Z. Lei, S. Z. Li, and X. Zou, "Selective refinement network for high performance face detection," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, 2019, pp. 8231–8238.

[55] G. Sun, X. Zhang, X. Jia, J. Ren, A. Zhang, Y. Yao, and H. Zhao, "Deep fusion of localized spectral features and multi-scale spatial features for effective classification of hyperspectral images," *International Journal of Applied Earth Observation and Geoinformation*, vol. 91, p. 102157, 2020.

[56] Z. Fang, J. Ren, S. Marshall, H. Zhao, Z. Wang, K. Huang, and B. Xiao, "Triple loss for hard face detection," *Neurocomputing*, 2020.

[57] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4700–4708.

[58] S. Gao, M.-M. Cheng, K. Zhao, X.-Y. Zhang, M.-H. Yang, and P. H. Torr, "Res2net: A new multi-scale backbone architecture," *IEEE transactions on pattern analysis and machine intelligence*, 2019.

[59] S. Belongie, J. Malik, and J. Puzicha, "Shape matching and object recognition using shape contexts," *IEEE transactions on pattern analysis and machine intelligence*, vol. 24, no. 4, pp. 509–522, 2002.

[60] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International journal of computer vision*, vol. 60, no. 2, pp. 91–110, 2004.

[61] P. Afshar, S. Heidarian, F. Naderkhani, A. Oikonomou, K. N. Plataniotis, and A. Mohammadi, "Covid-caps: A capsule network-based framework for identification of covid-19 cases from x-ray images," *arXiv preprint arXiv:2004.02696*, 2020.

[62] M. Rey-Area, E. Guirado, S. Tabik, and J. Ruiz-Hidalgo, "Fucit-net: Improving the generalization of deep learning networks by the fusion of learned class-inherent transformations," *arXiv preprint arXiv:2005.08235*, 2020.

[63] A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, Z. Lin, A. Desmaison, L. Antiga, and A. Lerer, "Automatic differentiation in pytorch," 2017.

[64] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[65] I. Loshchilov and F. Hutter, "Sgdr: Stochastic gradient descent with warm restarts," *arXiv preprint arXiv:1608.03983*, 2016.

[66] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 2117–2125.

[67] Z. Fang, J. Ren, S. Marshall, H. Zhao, S. Wang, and X. Li, "Topological optimization of the densenet with pretrained-weights inheritance and genetic channel selection," *Pattern Recognition*, vol. 109, p. 107608, 2021.

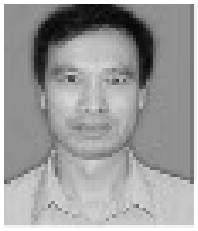
Zhenyu Fang received the B.Eng. degrees in Electronic and Electrical Engineering in 2016 from the University of Strathclyde Glasgow, Scotland (with first-class honours), and North China Electric Power University, Baoding, China. He received his Ph.D. in Electronic and Electrical Engineering at the University of Strathclyde in July 2020. His main interests are algorithm development for image classification, object detection and face detection.

Jinchang Ren received his B. E. degree in computer software, M.Eng. in image processing, D.Eng. in computer vision, all from Northwestern Polytechnical University, Xi'an, China. He was also awarded a Ph.D. in Electronic Imaging and Media Communication from the University of Bradford, Bradford, U.K.

Currently he is a Chair Professor of Computing Science, National Subsea Centre, Robert Gordon University, Aberdeen, UK. His research interests focus mainly on hyperspectral imaging, image processing, computer vision, big data analytics and machine learning. He has published over 300 peer reviewed journal/conferences papers, and acts as an Associate Editor for several international journals including IEEE TGRS and J. of the Franklin Institute. He is a Senior Member of the IEEE.

Calum MacLellan is a PhD candidate with the Dept. of biomedical Engineering, University of Strathclyde. His research interests include biomedical image analysis and deep learning.

Huihui Li received the Ph.D. degree in computer science and engineering from the South China University of Technology, Guangzhou, China. She is currently a Lecturer with the School of Computer Science, Guangdong Polytechnic Normal University. Her research interests include image processing, machine learning, and affective computing.



Huimin Zhao was born in Shaanxi, China, in 1966. He received the B.Sc. and the M.Sc. degree in signal processing in 1992 and 1997 from Northwestern Polytechnical University, Xian, China, respectively. He received the Ph.D. degree in electrical engineering from the Sun Yat-sen University in 2001.

He is currently a Professor and the Dean of School of Computer Sciences, Guangdong Polytechnic Normal University, Guangzhou, China. His research interests include image, video and information security technology.

Amir Hussain is founding Director of the Centre of AI and Data Science at Edinburgh Napier University. His research interests are cross-disciplinary and industry-led, aimed at developing cognitive data science and AI technologies, to engineer the smart and secure systems of tomorrow. He has (co)authored three international patents and over 400 publications, including 170+ international journal papers, 20 Books/monographs and over 100 Book chapters. He has led major national, EU and internationally funded projects and supervised over 30 PhD students. He is currently Chief Investigator for the COG-MHEAR programme grant funded under the EPSRC Transformative Healthcare Technologies 2050 Call. He is founding Editor-in-Chief of two leading international journals: Cognitive Computation (Springer Nature), and BMC Big Data Analytics. He is General Chair of IEEE WCCI 2020 (the world's largest technical event in Computational Intelligence), Vice-Chair of the Emergent Technologies Technical Committee of the IEEE Computational Intelligence Society (CIS), IEEE UK and Ireland Chapter Chair of the Industry Applications Society, (founding) Vice-Chair for the IEEE CIS Task Force on Intelligence Systems for e-Health, and a member of the UK Computing Research Committee (CRC).

Giancarlo Fortino (SM'12) is Full Professor of Computer Engineering at the Dept of Informatics, Modeling, Electronics, and Systems of the University of Calabria (Unical), Italy. He received a PhD in Computer Engineering from Unical in 2000. He is also distinguished professor at Wuhan University of Technology and Huazhong Agricultural University (China), high-end expert at HUST (China), senior research fellow at the ICAR-CNR Institute, and CAS PIFI visiting scientist at SIAT - Shenzhen. He is the director of the SPEME lab at Unical as well as co-chair of Joint labs on IoT established between Unical and WUT and SMU and HZAU chinese universities, respectively. His research interests include agent-based computing, wireless (body) sensor networks, and IoT.He is author of 430+ papers in int'l journals, conferences and books. He is (founding) series editor of IEEE Press Book Series on Human-Machine Systems and EiC of Springer Internet of Things series and AE of many int'l journals such as IEEE TAC,IEEE THMS,IEEE IoTJ,IEEE SJ,IEEE SMCM,Information Fusion,JNCA,EAAI, etc. He is cofounder and CEO of SenSysCal S.r.l., a Unical spinoff focused on innovative IoT systems. Fortino is currently member of the IEEE SMCS BoG and of the IEEE Press BoG, and chair of the IEEE SMCS Italian Chapter.